An Analysis of Amarillo's Yearly Rainfall Totals

By Robert Ashcraft
Retired Program Scientist - Pantex

Amarillo's local weather station has been measuring and reporting yearly rainfall totals since 1892. The driest year so far was 2011, with a total of 7.01 inches. The wettest year was 1923, with a total of 39.75 inches. The rainfall data, consisting of 125 values, can be found at the NOAA Online Weather Data website.

The purposes of this study were to (1) determine if the rainfall totals behaved like a random sample from a statistical distribution, (2) determine the type of the distribution, and (3) determine the distribution parameters that would provide the best approximation to the data.

The analysis began by generating a cumulative distribution for the data. To accomplish this, the data were sorted in ascending order. Then a quantile was computed for each data value using Hazen's formula,

$$q_i = \left(\frac{i - \frac{1}{2}}{n} \right)$$

where q_i is the i^{th} quantile, i ranges from 1 to n, and n is the number of data points. A plot was then made with the quantiles on the y-axis and the sorted rainfall totals on the x-axis, as shown in Figure 1.

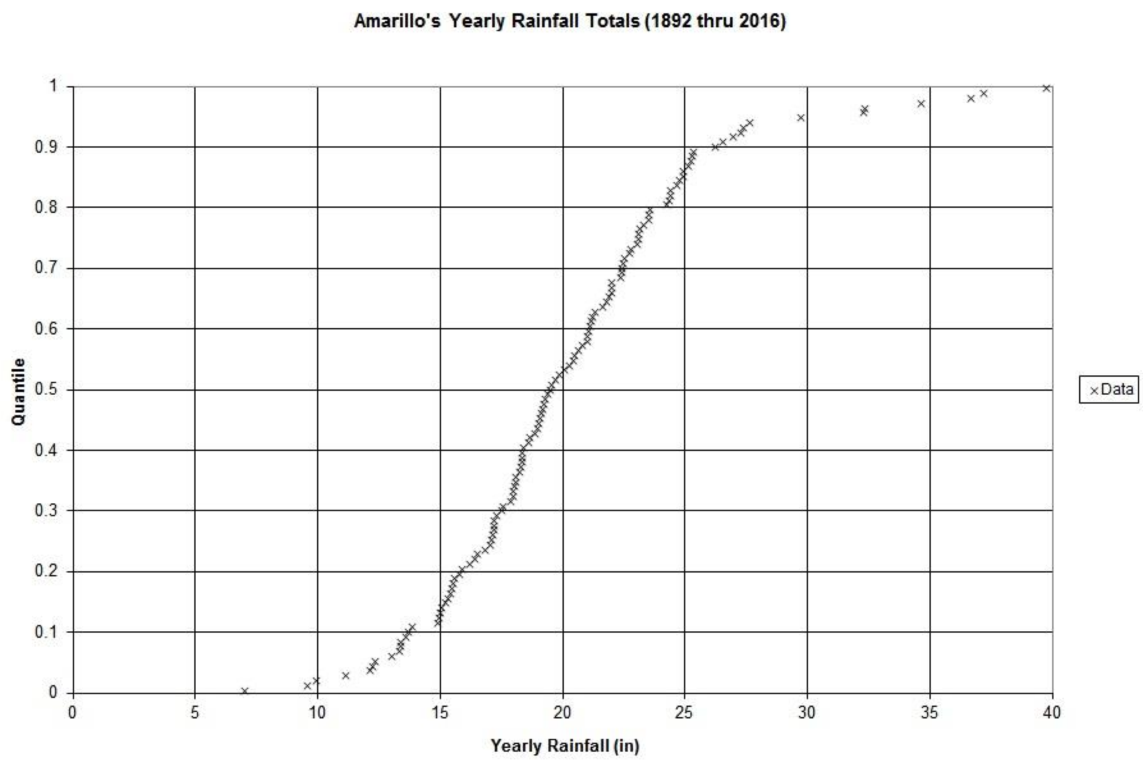

Figure 1. The cumulative distribution of Amarillo's yearly rainfall totals

The data appear to be roughly “S” shaped, but are not symmetric, since the right tail is longer than the left tail. Thus, we can rule out a normal distribution. Several asymmetric distributions were considered, such as log normal, square root normal, etc., but the distribution that fit the data the best was the logistic curve (see Appendix). Its equation is

$$F(x) = \frac{1}{1 + \exp\left(-\frac{\sqrt{x} - \alpha}{\beta}\right)}$$

where *F(x)* is the proportion of yearly rainfall totals that will be less than or equal to *x* inches, and α and β are parameters to be determined by a least squares regression.

The parameters α and β were found to be 4.448039161 and 0.319986248, respectively. The resulting curve is shown with the data in Figure 2, and the curve is an excellent approximation to the data.

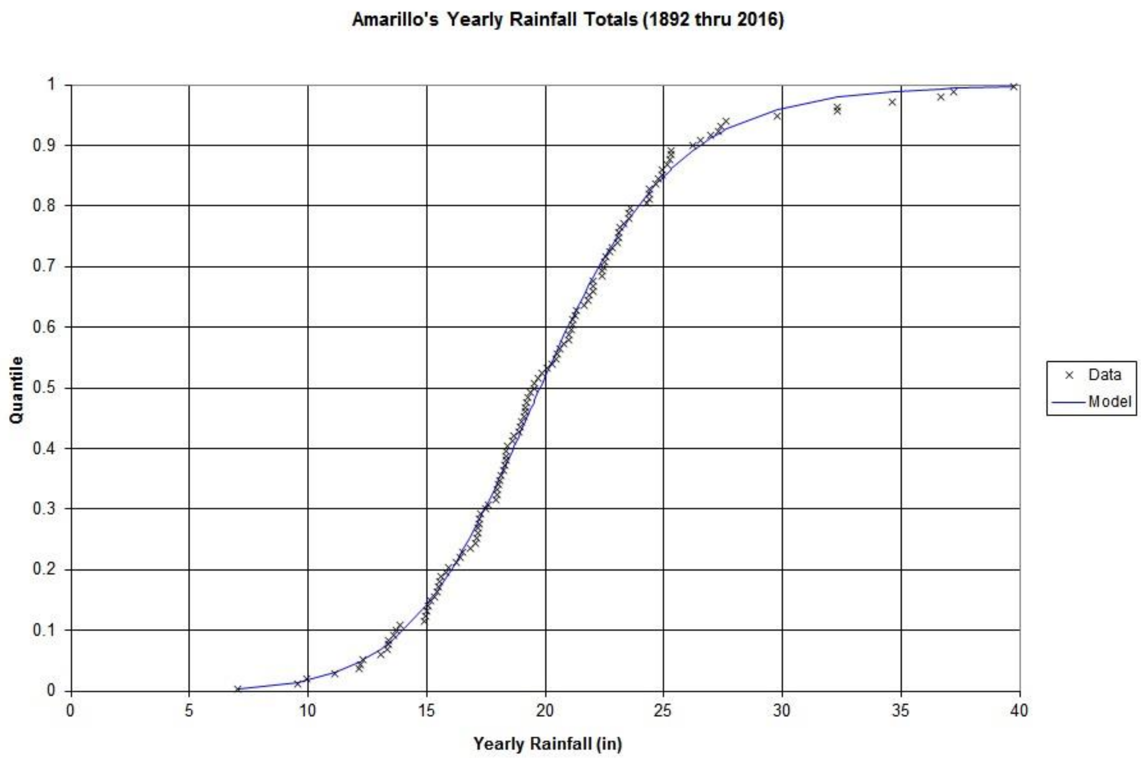

Figure 2. Amarillo’s yearly rainfall totals with the best fit logistic curve

Hence, for the past 125 years, Amarillo’s yearly rainfall totals have behaved like a random sample from a logistic distribution with parameters

α = 4.448039161
β = 0.319986248

Note that the median for the logistic curve is $\alpha^2 =$ 19.79 inches, which is a good approximation to the median of the rainfall data (19.52 inches).

APPENDIX

Hazen's Quantile Formula

There are several ways to calculate quantiles, but Hazen's formula is generally accepted in the analysis of hydrology data. Here is a link to MathWorld's article on quantiles with a list of commonly used formulas:

http://mathworld.wolfram.com/Quantile.html

The Logistic Curve

The equation for a logistic curve is

$$F(x) = \frac{1}{1 + \exp\left(-\frac{x-\alpha}{\beta}\right)}$$

In this equation, x would be the yearly rainfall total in inches. This curve did provide a good fit to the data, but a better fit was obtained by using the square root of the yearly rainfall total instead. Hence, the equation was

$$F(x) = \frac{1}{1 + \exp\left(-\frac{\sqrt{x}-\alpha}{\beta}\right)}$$

The inverse of this function is

$$x = \left[\alpha - \beta \ln\left(\frac{1}{F(x)} - 1\right)\right]^2$$

Examples

1. What proportion of yearly rainfall totals are less than or equal to 12 inches?

F(12) = 0.044153, or roughly 4.4%

2. What proportion of yearly rainfall totals range from 15 inches to 25 inches?

That will be F(25) – F(15) = 0.848766 – 0.142201 = 0.706565, or roughly 70.6%

3. 95% of yearly rainfall totals fall below what value?

F(x) = 0.95, so plug into the inverse function to get x = 29.04556 inches.